基于多模态图像信息及改进实例分割网络的肉牛体尺自动测量方法

翁 智[1,2], 范 琦[1,2], 郑志强[1,2*]

(1. 内蒙古大学 电子信息工程学院,内蒙古呼和浩特 010021,中国; 2. 省部共建草原家畜生殖调控与繁育国家重点实验室,内蒙古呼和浩特 010010,中国)

摘 要: [**目的/意义**]牛的体尺参数是反映牛身体发育状况的关键指标,也是牛选育过程的关键因素。为解决规模化肉牛牧场复杂环境对肉牛体尺的测量需求,设计了一种图像采集装置以及体尺自动测量算法。[**方法**]首先搭建肉牛行走通道,当肉牛通过通道后进入限制装置,用英特尔双目深度相机D455对牛只右侧图像进行RGB与深度图的采集。其次,为避免复杂环境背景的影响,提出一种改进后的实例分割网络Mask2former来对牛只二维图进行前景轮廓提取,对轮廓进行区间划分,利用计算曲率分析方法找到所需体尺测点。然后,将原始深度图转换为点云数据,对点云进行点云滤波、分割和深度图牛只区域的空值填充,以保留牛体区域的点云完整,从而找到所需测点并返回到二维数据中。最后,将二维像素点投影到三维点云中,利用相机参数计算出投影点的世界坐标,从而进行体尺的自动化计算,最终提取肉牛体高、十字部高、体斜长和管围4种体尺参数。[**结果与讨论**]改进的实例分割网络与Mask R-CNN、PointRend、Queryinst等模型相比具有更好的分割结果。采用本研究测得的这4种体尺平均相对误差分别为4.32%、3.71%、5.58%和6.25%。[**结论**]本研究开发的肉牛图像采集装置及相应的图像处理方法可以满足该牧场对肉牛体尺无接触自动测量误差小于8%的精度要求,为非接触式肉牛体尺自动化测量提供了理论与实践指导。

关键词: 肉牛体尺测量;深度学习;点云分割;实例分割;注意力机制;Mask2former

中图分类号: TP391.41;S823 **文献标志码:** A **文章编号:** SA202310007

引用格式:翁智, 范琦, 郑志强. 基于多模态图像信息及改进实例分割网络的肉牛体尺自动测量方法[J]. 智慧农业(中英文), 2024, 6(4): 64-75. DOI:10.12133/j.smartag.SA202310007

WENG Zhi, FAN Qi, ZHENG Zhiqiang. Automatic Measurement Method of Beef Cattle Body Size Based on Multimodal Image Information and Improved Instance Segmentation Network[J]. Smart Agriculture, 2024, 6(4): 64-75. DOI:10.12133/j.smartag.SA202310007 (in Chinese with English abstract)

0 引 言

2023年,中央一号文件《中共中央 国务院关于做好2023年全面推进乡村振兴重点工作的意见》要求大力推进种源等农业关键核心技术攻关,全面实施生物育种重大项目,扎实推进国家育种联合攻关和畜禽遗传改良计划。其中,牛的体况作为牛选育工作的参数尤为重要。牛是畜牧业发展的重要品种。准确测量牛的体尺数据对于探究其生长发育、繁殖性能具有重要意义,同时也是衡量其选育效果的重要依据[1]。

传统的牛体尺测量方法是采用测杖、卷尺和记录表通过人工将牛固定住进行测量。这种方法人工成本高,准确率低,容易造成牛的应激反应。随着人工智能的发展,越来越多的学者利用有关技术对牲畜体尺进行无接触式测量方法的研究。利用人工智能方法对牲畜体尺测量可分为基于二维数据和基于三维数据两种方法。

二维图像研究方法具有一定的局限性,当牛体背景较为复杂时,基于机器学习图像聚类的方法易受到环境影响,从而影响轮廓提取的精度[2]。Wang等[3]基于UNet网络对牦牛的前景进行提取,

收稿日期:2023-10-11

基金项目:国家自然科学基金项目(61966026);内蒙古自治区高等学校青年科技英才支持计划(NJYT23063);内蒙古自然科学基金项目(2021MS06014)

作者简介:翁 智,研究方向为人工智能与机器视觉。E-mail: wzhi@imu.edu.cn

*通信作者:郑志强,博士,副教授,研究方向为图像处理。E-mail: zqzheng@imu.edu.cn

利用双目相机来测得实际体尺数据，但是此分割方法并不适用于所有环境背景情况，并且在提取体尺关键测点的过程中是利用人工提取来实现的，缺少自动化过程。采用二维的方法对测点的提取有较大的局限性，不可能提取到非前景轮廓上的测点，从而造成体尺参数测量的局限性。

在三维数据领域，越来越多的研究者利用点云技术对牲畜体尺进行测量，利用深度相机采集牲畜的点云数据，通过点云分割，点云配准等技术，建立牲畜的轮廓三维数据，通过在点云数据上找测点，直接计算得到世界坐标系的点云坐标，从而计算体尺数据。Guo等[4]开发了一种基于三维点云数据的牲畜体长估计交互式软件LSSA_CAU，该软件提供了一套半自动加载、绘制、分割、姿态归一化、测量家畜全身体表点云数据的工具，但前提是假设牛或猪的形态相似且头朝前站立。Huang等[5]使用ShapeNet数据集的经典三维点云数据集（Point Cloud Data, PCD）训练Kd网络。通过光探测和测距（Light Detection and Ranging, LiDAR）传感器感知PCD的一系列过程，提取牛的轮廓，利用平均曲率和高斯曲率提取特征曲面形状的候选区域，在计算表面形状的特征直方图后，可以识别特征表面的中心，并最终计算出牛的身体尺寸。王可等[6]基于地面法向量矫正的主成分分析（Principal Component Analysis, PCA）方法，利用平面法矢量校正PCA主轴，建立了一种以旋转归一化为基础的猪体体尺测点提取方法。张婷和李琦[7]利用Kinect进行三维点云的重建，进而在点云图上进行曲线的拟合等操作来提取测点，实现体尺测量。Li等[8]利用Kinect相机从猪的正上方捕捉图像，从背表面点云中提取体尺寸参数。Yin等[9]通过对25头长白猪207组点云数据分析，提出了一种标准的姿态分类算法，将自动分类结果与基于猪标准姿态和非标准姿态的人工分类结果进行比较，对非标准姿态下猪体尺寸测量结果进行修正。Yang等[10]提出了一种基于智能手机的便携式奶牛身体非接触式三维测量系统。石炜等[11]利用四轴无人机进行图像采集，借助中值滤波降噪、图像三通道直方图均衡化等图像处理方法，对图像进行预处理，提高重建点云密度及测量精度，根据图像自带定位定姿系统（Position and Orientation System, POS）信息进行空中三角计算，获取平差确定点云的初始坐标数据之后进行三维重建，最后通过点云库（Point Cloud Library, PCL）点云处理中的随机采样一致性算法（RANdom SAmple Consensus, RANSAC）提取每头牛的点云，之后进行交互式测量。采用点云处理方法进行体尺测点自动化提取时，对来自不同视角的点云进行分割并配准出完整的待测对象的三维点云，只保留待测对象的点云数据。这对采集设备、采集环境、三维点云重建等算法与精度具有较高的要求，极易出现目标物体点云数据的缺失与细节不完整等问题。为解决上文提到的二维与三维方法的弊端，本研究采用二维、三维相结合的方法设计一种自动化非接触式肉牛体尺测量系统，主要工作如下。

1）由于采用机器学习进行图像分割对环境因素的鲁棒性不强，会导致分割难度较大，因此本研究改进深度学习实例分割网络Mask2former[12]，来获得更好的前景分割图像。

2）采用二维数据与三维数据结合的方法，针对不同测点位置，设计一种自动化体尺测点提取算法，利用Intel RealSense D455深度相机采集二维图片与深度数据，在二维图像上进行有关体尺测点的自动提取，在三维数据上对相关测点进行提取并返回到二维图像中与其他测点进行联系，最终再将所有体尺参数像素值投影到三维点云数据上，从而得到真实值。这种方法只需对点云数据进行一些简单必要的处理，保留有效信息即可。

3）搭建一种基于真实牧场环境下的测量平台，提出一种自动提取肉牛的体高、十字部高、体斜长和管围体尺参数的算法。

1 材料与方法

1.1 牛只尺寸的测量指标

本研究需要测量的牛体尺寸指标[13]如表1所示。待测体尺参数如图1所示。

表1 肉牛体尺测量标准

Table 1 Measurement standards for beef cattle body size

体尺参数	测量规定
体高	耆甲最高点到地面的垂直距离
体斜长	肩端到坐骨端后缘的直线距离
十字部高	腰角前缘对应的背部最高点至地面的垂直距离
管围	前肢胫部最细处的周长

1.2 设计牛体尺测量系统并获取人工测量的体尺数据

2023年7月15—22日，在内蒙古自治区赤峰市圣泉生态牧业有限公司采集数据，搭建采集通道和

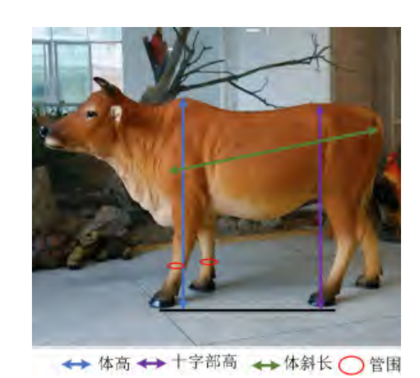


图1 牛体尺参数测量示意图

Fig. 1 Schematic diagram of measurement of beef cattle body size parameters

采集设备等实验环境。采集图像时，在距离肉牛右侧2～4 m处放置一台英特尔RealSense D455深度相机。采集图像的分辨率像素设置为848×480，帧率设置为30 f/s。采集环境的现场如图2所示。



a. 采集装置　　b. 采集示例

注：1. D455相机；2. 肉牛通过通道；3. 肉牛限制器。

图2 肉牛体尺参数实地采集环境图

Fig. 2 Collection environment of beef cattle body size parameters

如图3所示，在采集真实肉牛体尺数据时，测量人员使用测杖测量肉牛体尺。为减小误差，测量人员对每个体尺参数测量3次，取平均值作为最终人工测量的体尺数据。共对该牧场的137头牛进行相关体尺测量，所有体尺数值均以厘米（cm）为单位，保留小数点后两位小数。



a. 十字部高　　b. 体斜长　　c. 管围

图3 肉牛体尺参数人工测量照片

Fig. 3 Beef cattle body size measurement photos manually

肉牛体尺测量系统主要分为模型训练与测试、自动化体尺计算两部分。肉牛体尺测量系统方案设计如图4所示。

1.3 实例分割数据集构建

为对目标物体进行图像分割，增加模型的鲁棒

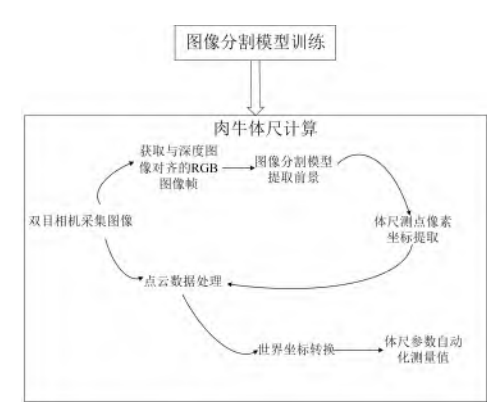


图4 肉牛体尺测量系统流程图

Fig. 4 Flowchart of body size measurement system of beef cattle

性，本研究构建了基于Microsoft Common Objects in Contex （MS COCO）数据集中牛的数据与该牧场采集数据的混合数据集，提取MS COCO数据集中2 055张牛的类别图，再将该牧场采集的视频数据集进行抽帧，选取580张图。为增强网络的鲁棒性和泛化能力，对图像进行翻转、镜像、缩放等手段进行数据增强，得到1 160张图像数据。数据分布情况见表2，数据集部分图像如图5所示。

表2 肉牛体尺实例分割实验数据集

Table 2 Instance segmentation experiment dataset for beef cattle body size

数据集	图像/张	手机/张	MS COCO/张
训练集	2 780	930	1 850
验证集	435	230	205
总共	3 215	1 160	2 055

数据集的标注精度直接影响检测结果。为减少标注可能带来的精度不足问题，本研究首先从官方COCO数据集的实例分割标注文件中将有牛的类别标注数据提取出来，然后使用官方百度旗下的Eiseg标注工具对自制数据集中的图像进行轮廓标注。为符合模型的训练要求，选择标注的数据集格式为JSON存储标注数据，其中包括目标边界框信息、目标面积、图像的名称、对应图像ID（Identity document）以及类别ID等信息，组成混合数据集。

2 研究方法

2.1 基于改进的Mask2former的牛体轮廓分割

Mask2former为所有图像分割任务训练了一个



a. 手机数据采集



b. 公开数据

图5　牛体实例分割试验部分数据集展示

Fig. 5 Partial dataset of instance segmentation of beef cattle body size

单一模型。三个图像分割任务（全景、实例和语义）在流行数据集上都取得了最佳效果，因此，在该算法上改进会获得更好的分割效果。

2.1.1　Mask2former结构原理

Mask2former是Facebook团队提出的基于Transformer的端到端的检测、分割框架，是对2021年maskformer[14]的进一步改进。

Maskformer的基本思想是把语义分割、实例分割用一个统一的框架、损失和训练过程来实现，都定义为掩膜分类任务，如果是语义分类任务在推理阶段加一个组合层即可获得预测结果。Maskformer与Deformable Detection Transformer[15]思想类似，都是去除人工经验，把实例分割作为一个集合预测问题来看待。首先用backbone提取图像基础特征，然后分2个分支：一个分支是通过pixel decoder模块生成pixel的embedding；另一个分支是通过transformer decoder生成N个queries，然后在queries后加全连接层和softmax就得到类别的概率估计。对于mask的预测，先通过N个queries经过两层的Multi-Layer Perceptron生成N个mask embedding，然后mask embedding和前面的pixel embedding相乘加sigmoid激活函数得到。在计算损失时用二分图匹配来配对预测和真值构造集合损失，从而简化pipeline实现端到端的目标检测。

Mask2former在maskformer的基础上主要做了3个方面的改进。一是Transformer Decoder中cross attention改成mask attention并和self attention换位置，以保证前景间互相attention忽略背景的干扰。二是增加多尺度特征，也就是把pixel decoder的不同层作为特征金字塔分别输入到transformer decoder的不同层中，以提升小目标的识别效果。三是在少量采样随机点而非全图上计算mask loss，以提升速度。本研究使用resnet50作为特征提取网络。在第0阶段，在骨干网络输出之前，分别加入注意力机制来对特征进行更好的表达。

2.1.2　注意力机制

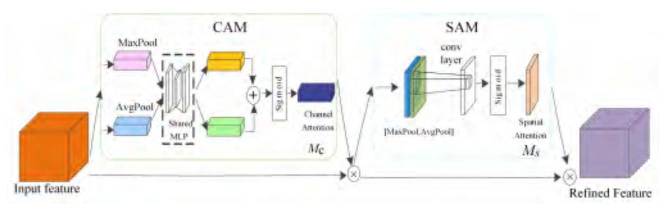
利用注意力机制，可以用高权重聚焦更重要的信息，用低权重忽略无关信息，而且权重还可以不断调整，从而在不同情况下也能选择重要信息。卷积块注意力模块（Convolutional Block Attention Module, CBAM）[16]能有效结合信道注意力和空间注意力，从两个方面加强关键特征的提取。坐标注意力模块（Coordinate Attention, CA）[17]不仅能获取信道间信息，还能考虑与方向相关的位置信息，帮助模型更好地定位和识别目标。注意力机制如图6所示。

在stage0中，只提取图像的低级特征，加入CBAM，可以为通道和空间特征分配权重，从而加强关键特征的提取。这部分改进命名为CB-Mask2former。在第4阶段末尾添加CA模块是为了在输出前对网络学习特征的表示进行最后的增强。这部分改进命名为CA-Mask2former。改进后的网络最终命名为CC-Mask2former，如图7所示。

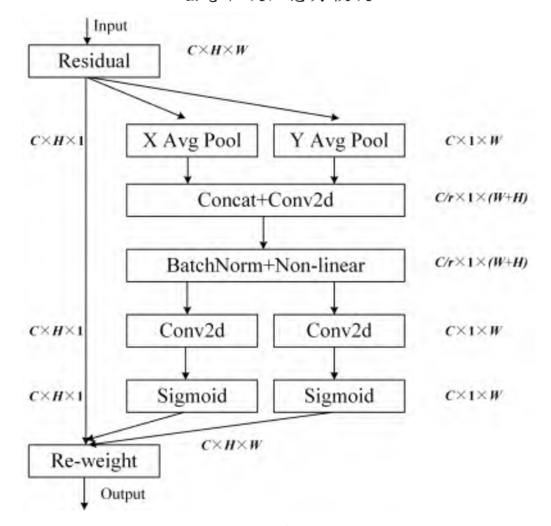
2.1.3　模型训练

实验所用服务器搭载英伟达RTX3090TI显卡，内存为64 GB。实验开发平台为64位Windows 10操作系统，12代i9处理器。软件平台为PyCharm，CUDA使用11.6版本，编程语言为Python 3.8.13，深度学习框架为Pytorch 1.12.0。

在模型训练时，将训练图像像素调整为512×512，epoch设置为60，batch size设置为12，由于batch size为12，因此训练集最终的迭代次数为13 900次。采用迁移学习方法，在训练开始时加入预训练模型。在训练中，优化器选用AdamW，学习率为0.000 1，权重衰减设为0.005。本研究采用平均精度均值（Mean Average Precision, mAP）作为评价指标。mAP是对每个类别的平均精度值



a. 卷积块注意力模块



b. 坐标注意力模块

图6 注意力机制结构图

Fig. 6 Structure of the attention mechanism

(Average Precision, AP)求和后的平均值。AP是以召回率(Recall, *R*)为横轴、精确率(Precision, *P*)为纵轴绘制P-R曲线并对其积分求出曲线下的面积而得到,即不同召回率下的精确率的平均值。*P*、*R*的计算如公式(1)和公式(2)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

式中:TP为实际正类预测为正类的数量;FP为实际负类预测为正类的数量;FN为实际正类预测为负类的数量。

2.2 三维数据处理

深度相机受到光照和周围环境噪声的影响,在二维图像中投影到点云深度图像时,可能会存在二维RGB图像中的待测像素点投影到深度值图像中对应像素坐标时出现该处为空值的情况,导致无法计算该点的三维坐标。因此,需要对深度图像进行优化来填补牛体区域空值的现象。由于深度相机采集到的深度值数据十分庞大,对原始深度数据进行优化会浪费大量的计算资源和时间,会对无关数据做很多无用功,因此在本小节中,对原始深度值数据进行处理,保留有效信息。

2.2.1 点云滤波

利用D455相机在对牧场中137头牛的深度图像进行采集后,每一张深度图像转成的点云有35万个点左右,数据量十分庞大。首先对原始点云数据进行直通滤波,因为牛身与其他的背景是沿着光心发射的方向,即Z轴的方向,所以可以根据相机与肉牛之间的距离,在Z轴方向进行阈值设置,去除Z值方向上大于5 m的深度值。直通滤波的计算如公式(3)所示。通过直通滤波后的点集$f(x, y, z)$只保留Z轴坐标在阈值z_1和z_2之间的点。

$$f(x, y, z) = \{(x, y, z) | z \in (z_1, z_2)\} \quad (3)$$

在经过直通滤波后会去除大量无用的点云数据,对直通滤波后的数据进行基于K近邻(K-Nearest Neighbor, KNN)算法的统计学滤波[18]。步

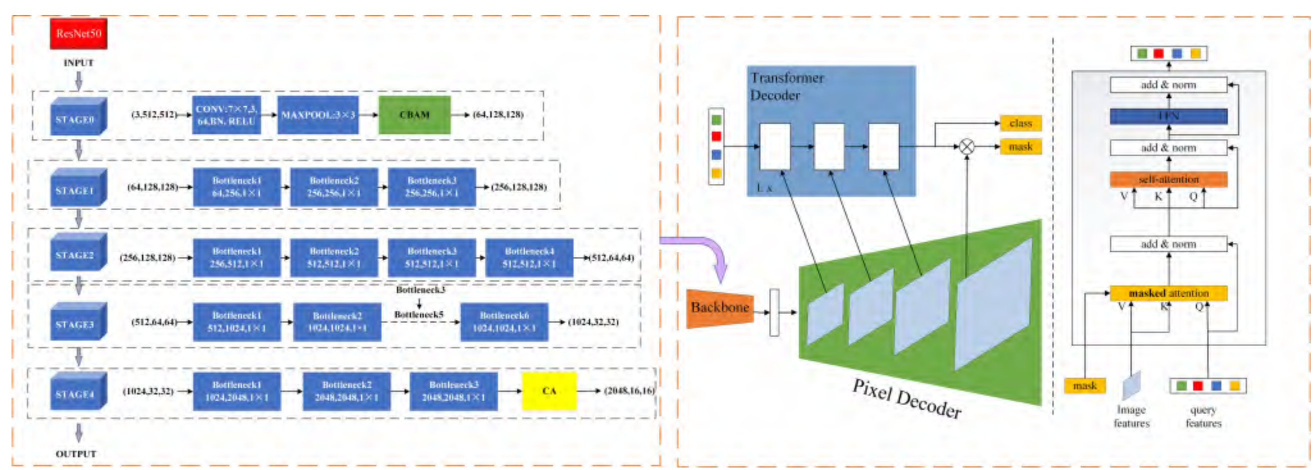


图7 CC-Mask2former模型结构图

Fig. 7 CC-Mask2former model structure

骤如下。

1）对于给定的点云数据集，首先选择一个固定的K值，即KNN的数量。

2）对于点云中的每个点，根据其K个最近邻点的统计特征，来判断该点是否为噪点。采用统计学中的离散度指标标准差来衡量点云的分布情况。标准差如公式（4）所示。

$$\sigma = \sqrt{\sum(x_i - u)^2/N} \tag{4}$$

式中：$\sum(x_i - u)^2$为所有数据与其平均值的差值的平方的总和；N为数据的数量；μ为数据的平均值。

3）如果一个点的K个最近邻点的离散度超过一个事先设置的阈值，那么认为该点是噪点，可以选择去除该噪点并进行后续的点云处理，或者保留该点。对于所有点重复上述过程，直到达到所有点的滤波要求。

2.2.2　点云分割

点云分割的目的是对目标点云进行更精细地提取。通过利用目标点云的空间、几何等特征，对原始点云进行分割，实现分割后的目标区域具有相同的特征。

本研究采用一种基于密度的空间聚类算法（Density-Based Spatial Clustering of Applications with Noise, DBSCAN）[19]，用于发现具有高密度的数据点的组成区域，并将其与低密度的区域分割开来。相比于传统的基于距离的聚类方式（如K均值聚类），DBSCAN能够有效处理具有不规则形状和不同密度的数据，并且可以自动确定簇的数量。算法实现步骤如下。

1）定义半径ε邻域（ε-Neighborhood），如公式（5）所示对于数据集D中的每个点p，它的半径ε邻域N（p）包含D中所有与p的距离小于等于ε的点。

$$N(p) = \{q|dist(p, q) \leqslant \varepsilon\} \tag{5}$$

2）定义核心点（Core Point）：如果点p的半径ε邻域中至少包含一定数量的点（称为密度阈值MinPts），则点p被认为是一个核心点。

3）定义直接密度可达（Directly Density-Reachable）：对于数据集D中的任意两个点p和q，如果q在p的半径ε邻域中，并且p是一个核心点，则称q从p直接密度可达。

4）定义密度可达（Density-Reachable）：对于数据集D中的任意两个点p和q，如果存在点序列p_1，p_2，…，p_n，其中p_1=p，p_n=q，且p_{i+1}从p_i直接密度可达，则称q从p密度可达。

5）定义密度相连（Density-Connected）：对于数据集D中的任意两个点p和q，如果存在一个点o，使得p和q都从o密度可达，则称p和q密度相连。最终将密度相连的点划分到同一个簇中。

2.2.3　深度值优化

将处理后的点云数据重新转换成新的深度图像。为防止存在空值的情况，在深度值为0的位置进行16领域的优化，使其范围扩大，将其16领域的深度值的平均值填充到该0位置上，可以将牛身体区域和牛前后腿之间的区域的空值进行合理的补充，从而保证RGB图像上的测点在对应像素位置的三维数据中不会出现空值，避免影响计算。16领域优化原理如图8所示，蓝色点为空值点，绿色区域为像素的16领域。

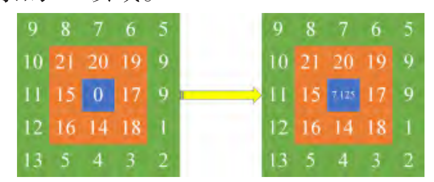


图8　深度值的16领域优化原理图

Fig. 8　Principle of 16-domain optimization for depth values

2.3 自动化提取体尺测点方法

为了便于体尺测点的提取，在本研究中需要在测点提取之前，统一将实例分割后的掩膜进行如下图像后处理操作：

1）去除分割图中的小对象部分，保留待测牛只掩膜的最大连通区域部分；

2）利用Canny算子进行边缘检测；

3）利用opencv对边缘检测图进行轮廓数据点的提取。

2.3.1　体高、十字部高参数提取方法

首先确定牛体轮廓的左上角(x_1, y_1)与右下角(x_2, y_2)，两点横纵坐标所包围的区域即为牛体轮廓所在区域。利用$x = [x_1 + (x_2 - x_1)/2]$纵垂线将牛体区域划分为左右两部分，由于在计算机图像中，坐标原点在图片左上角，因此分别在左右两部分区域找到各自区域的Y值最大的点来作为牛体的前后腿最低点，将两个点连起来的线段作为牛体尺测量的基准线。

耆甲点与十字部高测点提取流程如图9所示。

2.3.2　体斜长和管围提取方法

由于肩胛点处存在一个明显凸起，对于肩胛点

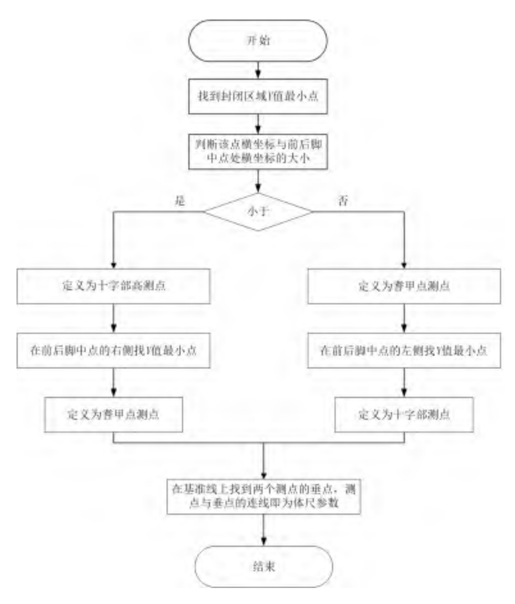


图9 肉牛体高、十字部高测点提取流程

Fig. 9 Point extraction process of beef cattle body height and cross section height measurement

的提取采用划分区间的方式，在找到耆甲点之后，将耆甲点的坐标作为区域的最小横纵坐标范围，将前脚点的纵坐标作为最大纵坐标范围，将轮廓的最右侧点的横坐标作为最大横坐标范围，在此区域内找到深度值矩阵中找到*Z*值最小处，并将该点的像素位置返回至RGB图像中，将该点作为肩胛点。

在找到肩胛点之后，将轮廓的最左侧点的横坐标作为最小横坐标范围，十字部高测点的横坐标作为最大横坐标范围，十字部高测点的纵坐标作为最小纵坐标范围，肩胛点的纵坐标作为最大纵坐标范围，对如图10所示红色区域所包围的黑色轮廓数据点进行最小二乘法拟合，找到曲率最大值的点定为坐骨端后缘测点。公式（6）为曲率计算公式。将肩端点到坐骨端后缘测点之间的距离定位体斜长。

$$k = y'' / \left\{ \left[1 + \left(y' \right)^2 \right]^{\frac{3}{2}} \right\} \tag{6}$$

将耆甲点的横坐标作为最小横坐标范围，轮廓的最大横坐标作为最大横坐标范围，前脚点的纵坐标作为最大纵坐标范围，从前脚点向上，找到前脚点与肩端点纵坐标差值的二分之一处，该处点的纵坐标作为最小纵坐标范围，如图10所示的绿色框区域。找出该区域包围的所有轮廓点，按纵轴方向自上而下绘制区域内宽度距离值变化曲线，计算同一纵坐标下的两个点的横坐标之间的像素距离，接着找到距离值曲线中基于多个连续相邻*y*方向上的像素点位置上，连续下降后、经过平稳一次后的第1个开始增大的点，这里的连续下降不包括相邻两个*y*像素点位置距离不变的情况，如图11所示。最后找到图10绿色区域中在该点*y*坐标值下的两个索引值，从而找到对应位置的两个*x*值，以两个点的*x*坐标值之间的距离作为直径，计算该直径下的圆的周长即为管围。



图10 坐骨端后缘测点区域和管围测点区域图示

Fig. 10 The region of the posterior border of the sciatic end of the point of measurement and the region of the tube perimeter measurement point

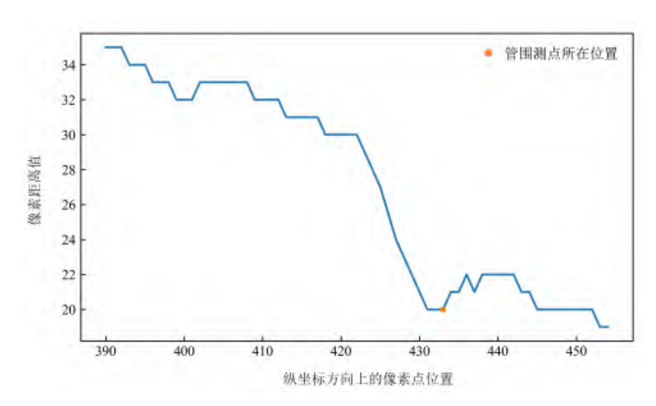


图11 管围测点区域的像素距离值曲线

Fig. 11 The pixel distance value curve of the tube circumference measurement area

2.4 投影点世界坐标计算

使用pyrealsense2库读出D455摄像机出厂时校准的摄像机参数，获取对齐后RGB光心参数以及校准后的深度帧的深度参数，如表3所示。

表3 Inter RealSense D455相机参数

Table 3 Inter RealSense D455 camera parameters

视角	光心坐标	焦距
深度图	（421.275，242.336）	（427.19，427.19）

点*P*的像素坐标为(u, v)，其三维投影点P_1在现实空间中的坐标为(X_w, Y_w, Z_w)则点*P*与P_1之间的变换关系见公式（7）。

$$\begin{cases} X_w = (u - u_0) \times z/f_x \\ Y_w = (v - v_0) \times z/f_y \\ Z_w = z \end{cases} \quad (7)$$

式中：z为深度信息；f_x、f_y分别为焦距信息；(u_0, v_0)为光心坐标。

3　结果与分析

3.1　图像分割实验结果

如图12所示，为$bbox_mAP_{0.5}$和$segm_mAP_{0.5}$在测试集上的60个epoch的曲线，满足测试精度。

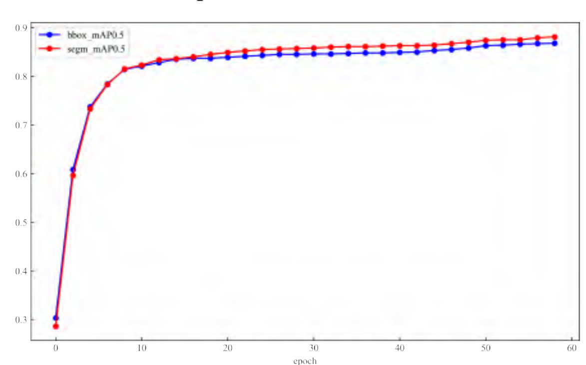


图12　CC-Mask2former模型牛体尺在测试结果

Fig. 12 Test results of beef cattle body of CC-Mask2former model

3.1.1　消融测试性能比较

为验证本研究提出的基于Mask2former的改进模型的有效性，分别在Mask2former、CB-Mask2former、CA-Mask2former和CC-Mask2former中进行实验，结果如表4所示。本研究提出的两部分改进措施都可以改善原始网络的指标。

表4　Mask2former添加注意力机制的消融实验对比

Table 4 Comparison of ablation experiments for Mask2former adding attention mechanisms

模型	$bbox_mAP_{0.5}$/%	$segm_mAP_{0.5}$/%
Mask2former	83.0	84.3
CB-Mask2former	83.5	84.9
CA-Mask2former	84.5	85.7
CC-Mask2former	86.8	88.1

3.1.2　不同模型的对比

为验证本研究提出的基于Mask2former的改进图像分割算法的有效性，分别在不同的模型上采用resnet50特征提取网络进行测试，与不同实例分割模型的对比结果如表5所示。

从表5可以看出，改进后的Mask2former与Mask R-CNN相比，$bbox_mAP_{0.5}$和$segm_mAP_{0.5}$分别提高2.6%和3.9%；与PointRend相比，$bbox_mAP_{0.5}$和$segm_mAP_{0.5}$分别提高16.4%和18.9%；与Queryinst相比，$bbox_mAP_{0.5}$和$segm_mAP_{0.5}$分别提高0.8%和3.9%。不同模型的图像分割结果示例如图13所示。

表5　不同分割模型与CC-Mask2former网络实验对比

Table 5 Experimental comparison of different segmentation models with CC-Mask2former network

模型	$bbox_mAP_{0.5}$/%	$segm_mAP_{0.5}$/%
Mask R-CNN[20]	84.2	84.2
PointRend[21]	70.4	69.2
Queryinst[22]	86.0	84.2
CC-Mask2former	86.8	88.1

3.2　三维数据处理结果分析

3.2.1　点云滤波

如图14所示，通过直通滤波后进行统计学滤波时，标准差阈值选为0.01，不同K值的选择直接影响最终点云的数量以及点云的滤波效果。

通过图14中各图的对比可知，K值越大滤波后的点云越少，牛身区域的点云存在丢失，并且耗费的计算时间越长。在本研究中K值选为5。

3.2.2　点云分割

对滤波后的点云进行DBSCAN算法的分割实验对比如图15所示。由此可知，在最小阈值密度一定的情况下，领域半径越小分割效果越好，在领域半径一定的情况下，最小密度阈值越大，分割后的点云数量越少，并且对牛体点云破坏较大。因此本研究选择使用最小密度阈值为5，领域半径为0.02进行点云分割。

3.2.3　深度值优化

对处理后的点云数据所对应的深度图进行优化，目的是填充牛体区域的空值点，分别对采用双线性插值、8领域平均值优化以及16领域平均值优化的结果进行对比，如图16所示，效果最好的是16领域平均值优化，它把牛身体部分和前后腿之间的空洞都很好地填补上。

3.3 体尺测点提取及体尺参数测量结果分析

3.3.1　体尺参数提取结果分析

采用本研究体尺参数提取方法对137头牛进行体尺参数提取。图17为个别牛只体尺参数提取结果，验证了本研究体尺参数提取算法的有效性。

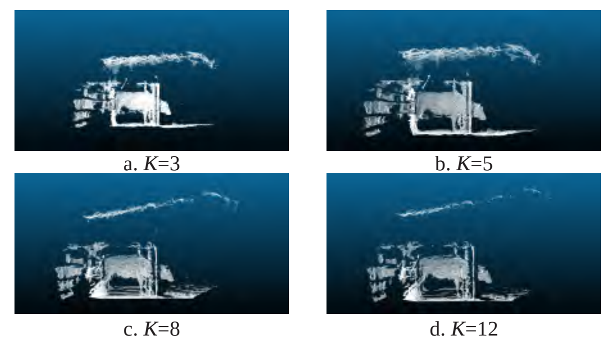
3.3.2　体尺测量结果与分析

对137头肉牛的体高、体斜长、十字部高和管

a. Mask R-CNN b. PointRend c. Queryinst d. CC-Mask2former

图13 不同实例分割模型的分割结果示例

Fig. 13 Examples of segmentation results for different instance segmentation models



a. K=3 b. K=5 c. K=8 d. K=12

图14 肉牛点云统计学滤波结果

Fig. 14 Statistical filtering results of beef cattle point clouds

围这4个体尺参数进行自动化测量。与人工测量值对比，计算不同体尺参数的相对误差，如图18所示。

体高的相对误差的平均值、最大值和最小值分别为4.32%、15%和1.16%，出现离群点是由于该牛只进入限制装置后出现前肢跪在地上的情况，导致出现了离群值，其余牛只正常进入装置后，由于不同姿势的影响会导致耆甲测点提取各有不同，属于正常现象。

十字部高的相对误差的平均值、最大值和最小值分别为3.71%、14.6%和0.68%，出现离群点是因为牛只偶尔排便时后腿一直弯曲所致。

体斜长的相对误差的平均值、最大值和最小值分别为5.58%、16.7%和1.14%，出现误差大的情况是受牛只排便等姿势的影响，坐骨端后缘测点提取出现偏差所致。

管围的相对误差的平均值、最大值和最小值分别为6.25%、14.9%和1.01%，出现误差大的原因是牛的不同站姿，当两条前腿重合度较大时，导致图像分割网络不能够同时分割出两条腿。

四种体尺参数平均相对误差最小的是十字部高。分析原因是由于十字部位置较为突出，牛只站姿的不同对十字部高的位置影响较小。

4 结 论

针对光照和复杂背景等因素的影响，为提高牛只前景图像的分割效果，本研究以实例分割模型

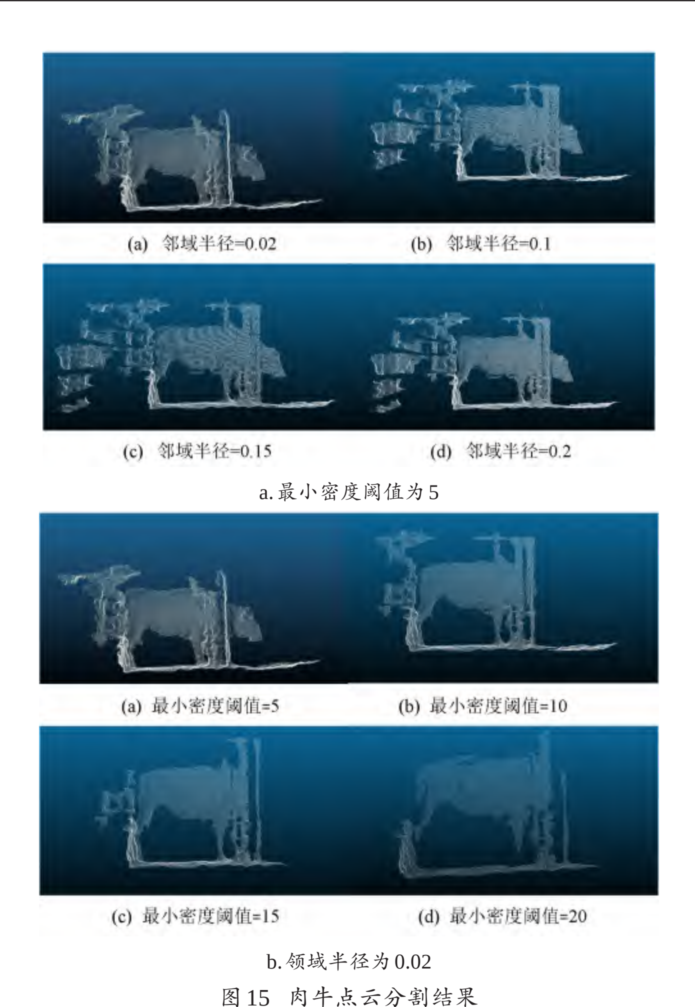


图15　肉牛点云分割结果

Fig. 15　Beef cattle point cloud segmentation results

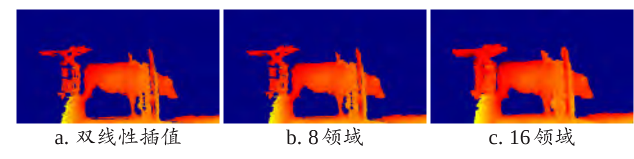


图16　肉牛点云分割后的深度图优化结果

Fig. 16　Depth map optimization results after point cloud segmentation of beef cattle

Mask2former为基础，加入CBAM模块和CA模块分别从不同角度提升模型对关键特征的提取能力。对比了经典的实例分割网络Mask R-CNN以及近年来几种新的实例分割网络，从分割精度以及最终分割的效果来看都有所提升，可以提取到较高精度、较为光滑的牛只前景图像。

设计一种二维与三维图像信息融合的方法，对137头牛只进行自动体尺测量。将4个体尺自动测量值与人工测量值进行对比，结果表明，体高、十字部高、体斜长和管围的平均相对误差分别为4.32%、3.71%、5.58%和6.25%。

本研究设计的体尺测量系统满足真实牧场环境需求，可较好地满足肉牛体尺自动化计算，为非接触式自动测量肉牛体尺提供了理论与实践指导。不足之处在于测量的体尺参数较少，测量周长类的体尺参数误差相对偏大。后期研究可以采用多视角方式增加待测体尺参数的数量，提高围度类参数精度。

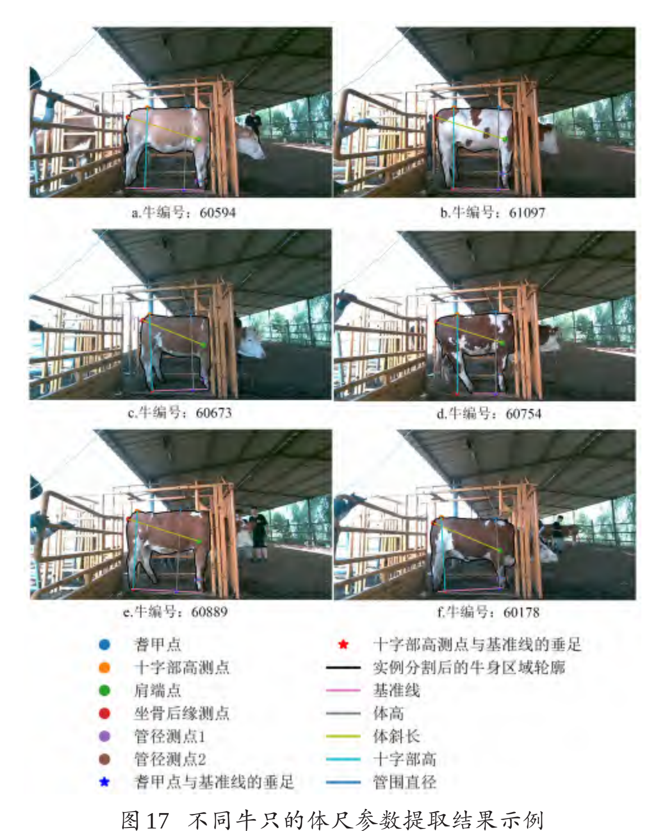
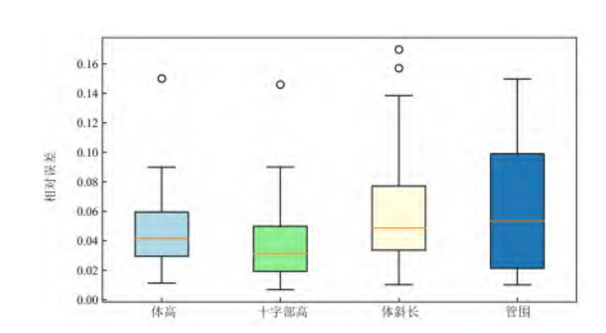


图17　不同牛只的体尺参数提取结果示例

Fig. 17　Example of extracted results of body size parameters for different cattles



注：箱形框内红色线代表相对误差率的均值，空心圆点代表误差离群点，横轴参数分别为体高、十字部高、体斜长和管围。

图18　肉牛体尺参数相对误差结果图

Fig. 18　Plot of relative error results for beef cattle body size parameters

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献:

[1] 张智慧, 李伟, 韩永胜. 牛体尺影响因素及其应用[J]. 中国畜牧杂志, 2018, 54(1): 9-13.
ZHANG Z H, LI W, HAN Y S. Influence factors and applications for body size of cattle[J]. Chinese journal of animal science, 2018, 54(1): 9-13.

[2] ZHANG A L, WU B P, TANA WUYUN C, et al. Algorithm of sheep body dimension measurement and its applications based on image analysis[J]. Computers and electronics in agriculture, 2018, 153: 33-45.

[3] WANG W, ZHANG Y, HE J, et al. Research on yak body ruler and weight measurement method based on deep learning and binocular vision[EB/OL]. Preprints, 2021, 2021120349. [2023-09-28]. https://doi. org/10.20944/preprints202112.0349.v2.

[4] GUO H, MA X D, MA Q, et al. LSSA_CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs[J]. Computers and electronics in agriculture, 2017, 138: 60-68.

[5] HUANG L W, GUO H, RAO Q Q, et al. Body dimension measurements of Qinchuan cattle with transfer learning from LiDAR sensing[J]. Sensors, 2019, 19(22): ID 5046.

[6] 王可, 郭浩, 刘威林, 等. 基于点云旋转归一化的猪体体尺测点提取方法[J]. 农业工程学报, 2017(S1): 253-259.
WANG K, GUO H, LIU W L, et al. Extraction of porcine body size measurement points based on point cloud rotational normalization[J]. Transactions of the Chinese society of agricultural engineering, 2017(S1): 253-259.

[7] 张婷, 李琦. 基于三维点云的牛体尺测量研究[J]. 黑龙江畜牧兽医, 2023, (3): 58-61.
ZHANG T, LI Q, et al. A study of bovine body measurements based on a three-dimensional point cloud[J]. Heilongjiang animal science and veterinary medicine, 2023, (3): 58-61.

[8] LI G X, LIU X L, MA Y F, et al. Body size measurement and live body weight estimation for pigs based on back surface point clouds[J]. Biosystems engineering, 2022, 218: 10-22.

[9] YIN L, ZHU J M, LIU C X, et al. Point cloud-based pig body size measurement featured by standard and non-standard postures[J]. Computers and electronics in agriculture, 2022, 199: ID 107135.

[10] YANG G Y, XU X S, SONG L, et al. Automated measurement of dairy cows body size via 3D point cloud data analysis[J]. Computers and electronics in agriculture, 2022, 200: ID 107218.

[11] 石炜, 陈禹州, 张万锴, 等. 基于无人机平台的牛体点云重构及体尺测量系统研究[J]. 畜牧与饲料科学, 2022, 43(4): 93-103.
SHI W, CHEN Y Z, ZHANG W K, et al. Cattle point cloud reconstruction and body size measurement system development based on unmanned air vehicle(UAV) platform[J]. Animal husbandry and feed science, 2022, 43(4): 93-103.

[12] CHENG B W, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2022: 1290-1299.

[13] 王永智. 牛的外貌鉴定、体尺测量及体重估测[J]. 养殖技术顾问, 2013(12): ID 20.
WANG Y Z. Appearance identification, body measurement and weight estimation of cattle[J]. Technical advisor for animal husbandry, 2013(12): ID 20.

[14] CHENG B W, SCHWING A G, KIRILLOV A. Per-pixel classification is not all you need for semantic segmentation[EB/OL]. arXiv: 2107.06278, 2021.

[15] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[EB/OL]. arXiv: 2010.04159, 2020.

[16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]// Computer vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

[17] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13713-13722.

[18] Abeywickrama T, Aamir Cheema M, Taniar D. k-Nearest Neighbors on Road Networks: A Journey in Experimentation and In-Memory Implementation[EB/OL]. arXiv: 1601.01549, 2016.

[19] Bäcklund H, Hedblom A, Neijman N. A density-based spatial clustering of application with noise[J]. Data Mining TNM033, 2011, 33: 11-30.

[20] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2961-2969.

[21] KIRILLOV A, WU Y X, HE K M, et al. PointRend: Image segmentation as rendering[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 9799-9808.

[22] FANG Y X, YANG S S, WANG X G, et al. Instances as queries[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2021: 6910-6919.

Automatic Measurement Method of Beef Cattle Body Size Based on Multimodal Image Information and Improved Instance Segmentation Network

WENG Zhi[1,2], FAN Qi[1,2], ZHENG Zhiqiang[1,2*]

(1. *School of Electronic Information Engineering, Inner Mongolia University, Huhhot* 010021, *China*;

2. *State Key Laboratory of Reproductive Regulation & Breeding of Grassland Livestock*, *Huhhot* 010010, *China*）

Abstract:

[Objective] The body size parameter of cattle is a key indicator reflecting the physical development of cattle, and is also a key factor in the cattle selection and breeding process. In order to solve the demand of measuring body size of beef cattle in the complex environment of large-scale beef cattle ranch, an image acquisition device and an automatic measurement algorithm of body size were designed.

[Methods] Firstly, the walking channel of the beef cattle was established, and when the beef cattle entered the restraining device through the channel, the RGB and depth maps of the image on the right side of the beef cattle were acquired using the Inter RealSense D455 camera. Secondly, in order to avoid the influence of the complex environmental background, an improved instance segmentation network based on Mask2former was proposed, adding CBAM module and CA module, respectively, to improve the model's ability to extract key features from different perspectives, extracting the foreground contour from the 2D image of the cattle, partitioning the contour, and comparing it with other segmentation algorithms, and using curvature calculation and other mathematical methods to find the required body size measurement points. Thirdly, in the processing of 3D data, in order to solve the problem that the pixel point to be measured in the 2D RGB image was null when it was projected to the corresponding pixel coordinates in the depth-valued image, resulting in the inability to calculate the 3D coordinates of the point, a series of processing was performed on the point cloud data, and a suitable point cloud filtering and point cloud segmentation algorithm was selected to effectively retain the point cloud data of the region of the cattle's body to be measured, and then the depth map was 16. Then the depth map was filled with nulls in the field to retain the integrity of the point cloud in the cattle body region, so that the required measurement points could be found and the 2D data could be returned. Finally, an extraction algorithm was designed to combine 2D and 3D data to project the extracted 2D pixel points into a 3D point cloud, and the camera parameters were used to calculate the world coordinates of the projected points, thus automatically calculating the body measurements of the beef cattle.

[Results and Discussions] Firstly, in the part of instance segmentation, compared with the classical Mask R-CNN and the recent instance segmentation networks PointRend and Queryinst, the improved network could extract higher precision and smoother foreground images of cattles in terms of segmentation accuracy and segmentation effect, no matter it was for the case of occlusion or for the case of multiple cattles. Secondly, in three-dimensional data processing, the method proposed in the study could effectively extract the three-dimensional data of the target area. Thirdly, the measurement error of body size was analysed, among the four body size measurement parameters, the smallest average relative error was the height of the cross section, which was due to the more prominent position of the cross section, and the different standing positions of the cattle have less influence on the position of the cross section, and the largest average relative error was the pipe circumference, which was due to the influence of the greater overlap of the two front legs, and the higher requirements for the standing position. Finally, automatic body measurements were carried out on 137 beef cattle in the ranch, and the automatic measurements of the four body measurements parameters were compared with the manual measurements, and the results showed that the average relative errors of body height, cross section height, body slant length, and tube girth were 4.32%, 3.71%, 5.58% and 6.25%, respectively, which met the needs of the ranch. The shortcomings were that fewer body-size parameters were measured, and the error of measuring circumference-type body-size parameters was relatively large. Later studies could use a multi-view approach to increase the number of body rule parameters to be measured and improve the accuracy of the parameters in the circumference category.

[Conclusions] The article designed an automatic measurement method based on two-dimensional and three-dimensional contactless body measurements of beef cattle. Moreover, the innovatively proposed method of measuring tube girth has higher accuracy and better implementation compared with the current research on body measurements in beef cattle. The relative average errors of the four body tape parameters meet the needs of pasture measurements and provide theoretical and practical guidance for the automatic measurement of body tape in beef cattle.

Key words: cattle body size measurement; deep learning; point cloud segmentation; instance segmentation; attention mechanism; Mask2former

Foundation items: National Natural Science Foundation of China (61966026); Program For Young Talents of Science And Technology In Universities of Inner Mongolia Autonomous Region Grant (NJYT23063); Natural Science Foundation of Inner Mongolia under Grant (2021MS06014)

Biography: WENG Zhi, E-mail: wzhi@imu.edu.cn

***Corresponding author:** ZHENG Zhiqiang, E-mail: zqzheng@imu.edu.cn